

## **Supplementary Information**

***In Silico* Identification, Characterization and Diversity Analysis of RNAi  
Pathway Gene Families in Sweet Orange (*Citrus sinensis* L)**

**Supplementary Table S1: Gene location in different scaffold.**

Gene Name	Accession Number	location	Location in scaffolds (Phytozome)												
<b><i>CsDCLs</i></b>															
CsDCL1	orange1.1g000174m	scaffold00001:3480331..3490933	0	500,000	1,000,000	1,500,000	2,000,000	2,500,000	3,000,000	3,500,000	4,000,000	4,500,000	5,000,000	5,500,000	
CsDCL2	orange1.1g000607m	scaffold00367:74912..87639	0	20,000	40,000	60,000	80,000	100,000	120,000	140,000	160,000				
CsDCL3	orange1.1g000379m	scaffold00013:998934..1009643	0	100,000	200,000	300,000	400,000	500,000	600,000	700,000	800,000	900,000	1,000,000	1,100,000	
CsDCL4	orange1.1g000380m	scaffold00068:219006..230745	0	50,000	100,000	150,000	200,000	250,000	300,000	350,000	400,000	450,000	500,000	550,000	600,000
<b><i>CsAGOs</i></b>															
CcAGO1	orange1.1g001466m	scaffold00674:51962..59926	0	5,000	10,000	15,000	20,000	25,000	30,000	35,000	40,000	45,000	50,000	55,000	60,000
CcAGO4	orange1.1g002449m	scaffold00028:907375..916121	0	100,000	200,000	300,000	400,000	500,000	600,000	700,000	800,000	900,000	1,000,000	1,100,000	
CcAGO5a	orange1.1g002204m	scaffold00595:85931..92382	0	5,000	10,000	15,000	20,000	25,000	30,000	35,000	40,000	45,000	50,000	55,000	60,000
CcAGO5b	orange1.1g037086m	scaffold00595:96826..99593	0	10,000	20,000	30,000	40,000	50,000	60,000	70,000	80,000	90,000			
CcAGO5c	orange1.1g003630m	scaffold03700:12..6517	0	500	1,000	1,500	2,000	2,500	3,000	3,500	4,000	4,500	5,000	5,500	6,000
CcAGO6	orange1.1g002661m	scaffold00067:338566..346634	0	50,000	100,000	150,000	200,000	250,000	300,000	350,000	400,000	450,000	500,000	550,000	600,000
CcAGO7	orange1.1g001684m	scaffold00003:3309646..3313553	0	500,000	1,000,000	1,500,000	2,000,000	2,500,000	3,000,000	3,500,000	4,000,000				
CcAGO10	orange1.1g001954m	scaffold00011:54251..63917	0	200,000	400,000	600,000	800,000	1,000,000	1,200,000	1,400,000	1,600,000	1,800,000	2,000,000	2,200,000	
<b><i>CsRDRs</i></b>															
CsRDR1	orange1.1g002586m	scaffold00058:231382..236081	0	50,000	100,000	150,000	200,000	250,000	300,000	350,000	400,000	450,000	500,000	550,000	600,000
CsRDR2	orange1.1g001183m	scaffold00020:1255446..1250927	0	200,000	400,000	600,000	800,000	1,000,000	1,200,000						
CsRDR3	orange1.1g001771m	scaffold00027:638984..650509	0	100,000	200,000	300,000	400,000	500,000	600,000	700,000	800,000	900,000	1,000,000	1,100,000	
CsRDR6	orange1.1g041430m	Scaffold d00051:398116..402488	0	50,000	100,000	150,000	200,000	250,000	300,000	350,000	400,000	450,000	500,000	550,000	600,000

**Supplementary Table S2:** Sub-cellular Localization of the predicted proteins

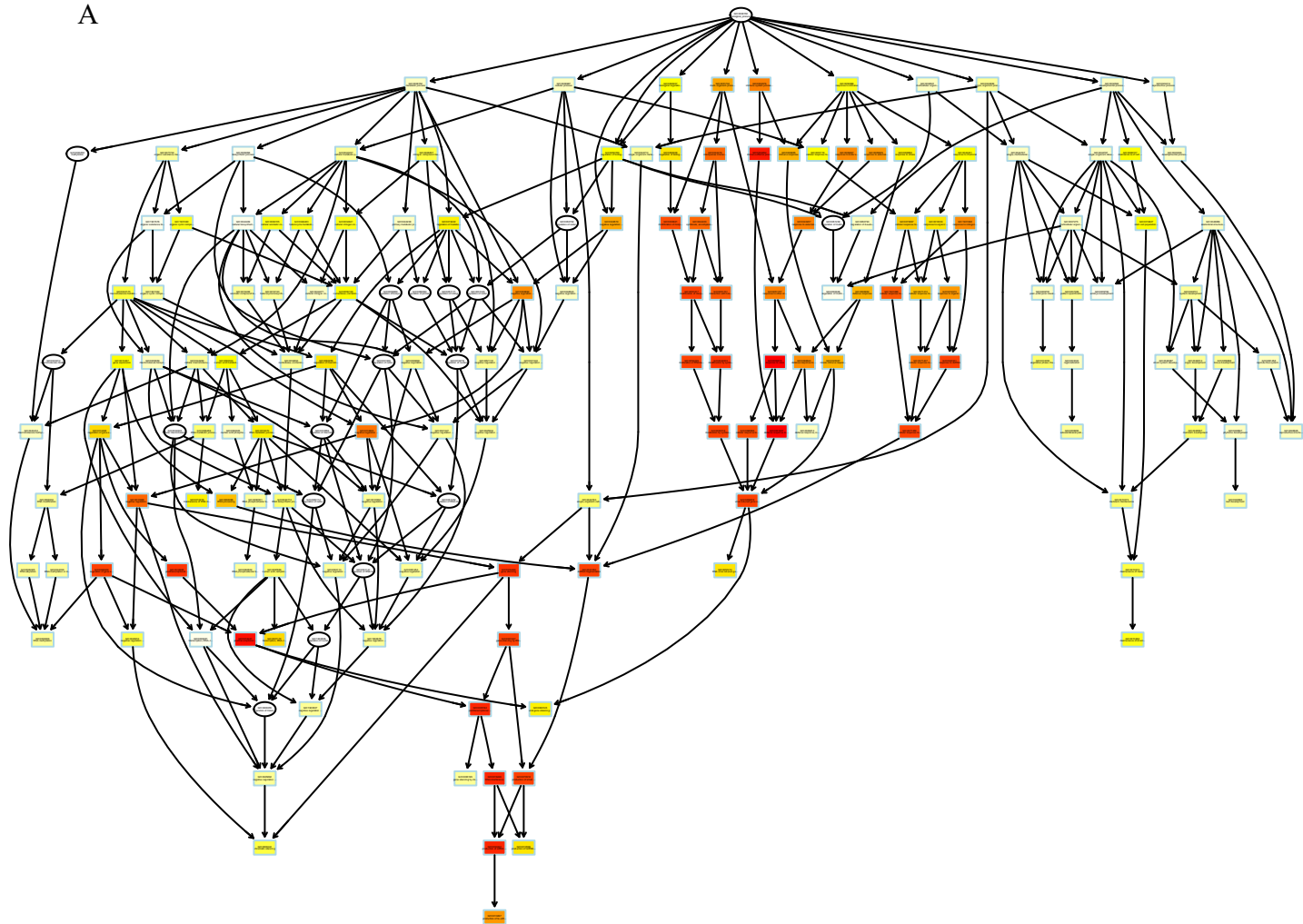
	extra	cytos	membr	ER	mito	golgi	plast	nucl	vacu	pero
<b>CsDCL</b>	0	100	25	0	0	0	0	0	0	0
<b>CsAGO</b>	0	100	0	0	12.5	0	0	50	0	0
<b>CsRDR</b>	0	100	25	0	0	0	25	25	0	0

cytosol (cytos), endoplasmic reticulum (ER), extracellular (extra), golgi apparatus (golgi), membrane (membr), mitochondria (mito), nuclear (nucl), peroxisome (pero), plastid (plast) and vacuole (vacu).

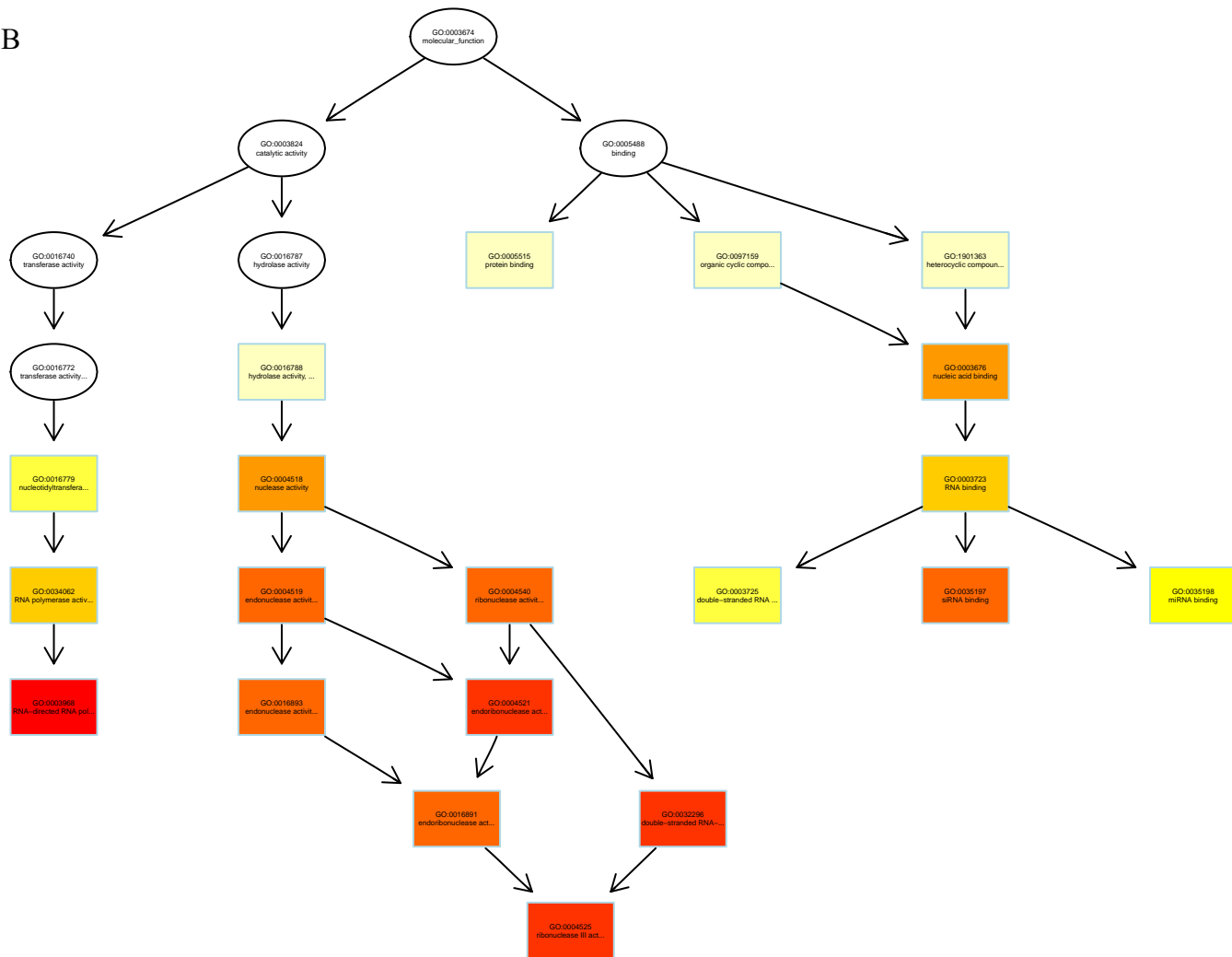
**Supplementary Table S3:** Distribution of TF families those regulating RNAi genes.

TF Family	Count	Percent (%)
ERF	29	21.16788
NAC	20	14.59854
WRKY	20	14.59854
bZIP	10	7.29927
Dof	6	4.379562
MYB	6	4.379562
TCP	6	4.379562
bHLH	5	3.649635
LBD	5	3.649635
GATA	4	2.919708
B3	3	2.189781
BES1	3	2.189781
AP2	2	1.459854
BBR-BPC	2	1.459854
C2H2	2	1.459854
EIL	2	1.459854
SBP	2	1.459854
ARF	1	0.729927
CAMTA	1	0.729927
GRAS	1	0.729927
HD-ZIP	1	0.729927
MIKC_MADS	1	0.729927
Nin-like	1	0.729927
TALE	1	0.729927
Trihelix	1	0.729927
WOX	1	0.729927
ZF-HD	1	0.729927

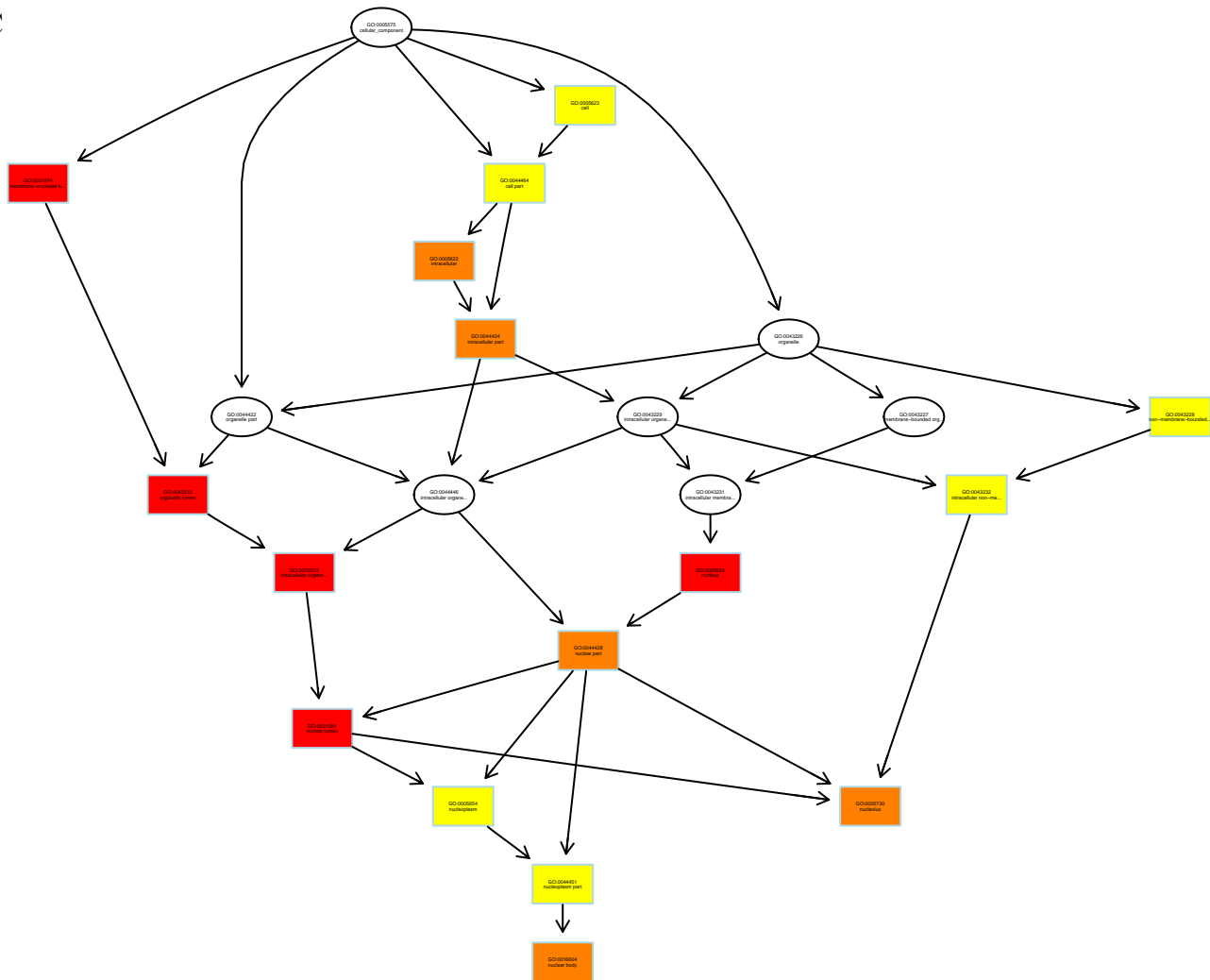
A



B



C



**Supplementary Fig. S1:** GO enrichment analysis of the predicted RNAi genes A) biological process, B) molecular function and C) cellular process. In the directed acyclic graph (DAG) the downstream term corresponds to a subset of the upstream term. The significant ( $P < 0.05$ ,  $FDR < 0.05$ ) GO terms are in colored boxes (the degree of color saturation is positively correlated to the enrichment level of the GO term), and non-significant terms are in white boxes.



	AP2	ARF	B3	BBR-BPC	BES1	bHLH	bZIP	C2H2	CAMTA	Dof	EIL	ERF	GATA	GRAS	HD-ZIP	LBD	MIKC_MADS	MYB	NAC	Nin-like	SBP	TALE	TCP	Trihelix	WOX	WRKY	ZF-HD
CsDCL1	1		1		2		3	1	1	4			1			1	1		1							20	
CsDCL2											2								5								1
CsDCL3	1				2	3	1	2		4				1			1										
CsDCL4					1							5							3		1	1	6	1		4	
CsAGO1					1	5	10												17								
CsAGO4	1																				1						
CsAGO5a	1				2					6		28	1	1		4	1	2				1					
CsAGO5b								1				2													1		
CsAGO5c								1				10				3	1			1							
CsAGO6								1						1	1												
CsAGO7	1				1			1		3		1						1	2	7							
CsAGO10	1		1	2				1		4			3					1	2								
CsRDR1							2																				
CsRDR2												2											5			3	
CsRDR3			1	1													1										
CsRDR6												1															

**Supplementary Fig. S2:** Distribution of TF families corresponding to genes. Rows of the figure represent the predicted RNAi genes and the columns represent the families of the TFs. The number indicates the TF families regulate the RNAi genes.